

RESEARCH

Open Access

Distributed video coding with block mode decision to reduce temporal flickering

Chang-Ming Lee^{1,2*}, Zhi-Heng Chiang¹, Dung-Chan Tsai² and Wen-Nung Lie^{2,3}

Abstract

The common distributed video coding (DVC) systems treat input video frames group by group. In each group of pictures (GOP), usually the first frame, called key frame, is intra-coded and the others are Wyner-Ziv (WZ)-coded. This GOP coding structure presents a fixed and inefficient coding mode switch (key or WZ) at group boundaries, thus preventing a DVC system from adapting the rate-distortion (R-D) performance to varying video content. In this work, a structure of temporal group of blocks (TGOB) with dynamic coding mode (key/WZ) decision is presented. This dynamic TGOB coding architecture determines each image block to be a key block or a WZ block based on spatiotemporal image analysis, resulting in a mode switch of fine granularity in both the spatial and temporal domains. As a consequence, not only the overall coding efficiency is improved, but also the temporal flickering artifacts for the reconstructed video are reduced substantially. Experimental results show that our proposed DVC scheme with block mode decision achieves up to 2.85 dB of quality gain and also up to 51% of temporal flickering reduction, compared to the well-known DISCOVER system.

Keywords: Distributed video coding; Wyner-Ziv theorem; Low-density parity-check accumulate codes; Flickering

1 Introduction

In the past, the development of video coding techniques aimed at the broadcasting applications (i.e., one transmitter and a plurality of receivers). Hence, a fast and low-cost design of a video decoder at the receiver part is emphasized. This implies that the encoder at the broadcast transmitter, which is in charge of the coding efficiency, might be subject to high computing complexity. However, this heavy-encoder light-decoder architecture is inappropriate for emerging applications, such as video surveillance systems or wireless sensor networks, where distributed video sensors are responsible for collecting video information there around and transferring them to a centralized controller. In those cases, the requirement of designing a low-complexity video encoder and a high-complexity video decoder overthrows the traditional concept. This encourages the development of distributed video coding (DVC) [1] that is based on the famous Slepian-Wolf (SW) [2] and Wyner-Ziv (WZ) [3] theorems.

The principal idea behind DVC is to encode correlated sources X and Y separately, while to decode them jointly with a decoded source (denoted as \hat{Y}) being probably processed and then considered in decoding the other source (i.e., X) as the side information (denoted as X'). The decoder for the source X differs from the traditional one (e.g., H.264/AVC video decoder) in an aspect that X is channel-coded and parity bit information of X is transmitted to the receiver for error decoding with the side information X' . In this way, the dominant computational loading is shifted from the encoder to the decoder. Although many applications of distributed source coding (DSC) seem practical, most DVC systems suffer from degraded coding efficiency compared to H.264/AVC. Currently, many researches are devoted to improving the coding performance of DVC systems. For example, to fulfill WZ decoding, a statistical correlation noise modeling (CNM) [4], also considered as a virtual channel, is proposed to describe the dependency between a source and its side information in WZ coding. A DVC decoder usually contains a function module to generate side information of X from the other decoded source \hat{Y} . Basically, the efficiency of WZ coding is dominated by the accuracy of both the generated side information and the CNM. The

* Correspondence: cmlee@ccu.edu.tw

¹Department of Communications Engineering, National Chung Cheng University, Chia-Yi 621, Taiwan

²Department of Electrical Engineering, National Chung Cheng University, Chia-Yi 621, Taiwan

Full list of author information is available at the end of the article

more accurate these two estimates are, the fewer parity bits are needed to reconstruct X and better rate-distortion (R-D) performance can be achieved. Many research works are dedicated to generate accurate side information. The motion-compensated temporal interpolation (MCTI) [5,6] is a sophisticated technique to produce the side information at the DVC decoder, and it would result in a reduced amount of parity information requested by the decoder through a feedback channel. Moreover, an improved fusion of global and local side information at the decoder side is also proposed in [7]. The review of several important side information generation methods can be found in [8]. Additionally, new applications of DVC in [9,10] are inspired with these investigations.

Typically, WZ coding is mainly realized by channel codes with capacity near the Shannon limit such as turbo codes [11,12] and low-density parity-check (LDPC) codes [13]. In a practical DVC codec, the structure of channel codes needs to be adapted to the varying statistics in the CNM. The first work about LDPC-based rate-adaptive codes which can be utilized in DVC is the construction of low-density parity-check accumulate (LDPCA) codes [14], in which a number of accumulated bits is sent to the decoder to accomplish the WZ decoding according to varying correlation noises. The corresponding LDPC-based systems outperform the turbo-based schemes in terms of compression performance. However, several techniques, considering flexibility and granularity of coding units (e.g., a hybrid block-frequency architecture [15]), implement WZ coding with block-based turbo codes. This work, though adopting the LDPCA code which has been used in many frame-based DVC schemes, however, determines coding modes on block basis (instead of frame basis). This is advantageous for maintaining good coding performance on one hand and reducing flickering artifact on the other hand, which will be discussed later in this paper.

Generally, the encoder lacks *a priori* knowledge about the number of parity bits required at the decoder to reconstruct X . Hence, parity bits are generated in batches, partitioned into packets, and stored in encoder buffers. Based on the side information X' generated from \hat{Y} , the decoder requests parity bits packet by packet via a feedback channel until the success of channel decoding of X . This indicates a delay, and channel decoding is performed iteratively with newly requested parity bits. Such procedures inevitably result in a huge computing load at the decoder. Several works suggest realizing bit rate estimation at the encoder [16-19]. A rate estimation based on the residue between the current frame and the side information is proposed [16], where the parity bits are allocated dynamically according to the motion activity. An entropy-based encoder rate estimation in transform domain is also proposed [17], where the Laplacian correlation model is used to estimate the entropy. Robustness of

the rate estimation is improved by the use of an error detection method based on cyclic redundancy checksum (CRC).

Temporal smoothness plays an important role for the human visual system in evaluating visual quality of a video. A decoded video sequence with lower flicker is more preferable, whether for distributed or conventional video coding systems. Existing DVC systems [18,19] mostly divide the whole video sequence into predefined groups of pictures (GOPs) for encoding, where the GOP size is fixed and frames in a GOP are organized as one intra-coded key frame (denoted as KF) followed by others which are WZ-coded. In general, KFs are intra-encoded based on a MPEG- or H.264/AVC-like standard, and WZ frames (denoted as WZFs) are based on channel coding techniques (e.g., LDPCA or turbo codes). To reconstruct WZFs at the decoder, decoded KFs act as the sources for side information generation (via, e.g., interpolation techniques). In general, temporal correlation between consecutive frames may be substantially high for static scenes and low for dynamic ones. Fixed and improper selection of the coding modes (KF or WZF) in a sequence incurs not only temporal flickering artifacts, but also degradation of the coding efficiency. Thus, the frame-based DVC schemes usually suffer from noticeable temporal fluctuation due to image quality unbalance caused by different coding methods between KFs and WZFs. Traditional KF/WZF coding schemes also cannot adapt to frames of varying content. Therefore, the fixed GOP structure is not nimble for varying video content, and an adaptive framework to overcome the above shortcoming is imperative. To deal with this situation, block-based, instead of traditional frame-based, WZ video coding schemes are proposed [20,21]. In our preliminary work on a pixel-domain DVC system [21], each block in a frame is categorized into a key block (denoted as KB) or a WZ block (denoted as WZB) after taking into consideration its temporal and spatial statistics. A better coding performance compared to conventional frame-based DVC schemes can be achieved, owing to the dynamically spatial key block decision adapting to varying content in each frame. Another advantage is to provide enhanced temporal consistency owing to the dynamically temporal key block decision adapting to varying content along the sequence. That is, finer granularities in both the spatial and temporal directions can be achieved in determining the coding modes. In this paper, the prior DVC scheme with block mode decision is further extended to the transform domain, which ensures better temporal consistency and hence further improves visual experience during the playback of the reconstructed video. Some comparisons with the distributed coding for video services (DISCOVER) codec [22], as well as a more complete and rigorous performance analysis of the proposed scheme, are addressed in this paper.

The rest of this paper is organized as follows. Section 2 overviews DVC systems based on flexible GOP structures in the literature. Section 3 provides details about the proposed framework which is based on the concept of temporal group of blocks (TGOB). Section 4 presents our experiment results with comparisons to the DISCOVER codec [22] in R-D performance and resulting temporal flickering artifacts. Finally, Section 5 draws a conclusion and future work.

2 DVC system with flexible GOP structure

In conventional video codecs (e.g., MPEG-X/H.26X/H.264), a GOP structure is predefined to consist of an intra-coded frame and several inter-coded frames that follow. The main advantages of such a GOP structure are to avoid the propagation of channel errors due to motion compensation and to offer random access of each distinct GOP. However, this kind of fixed GOP structure is not suitable for DVC systems that aim to provide a simple encoder, where exhaustive exploration of temporal redundancy is not allowed.

2.1 Conventional frame-based DVC scheme

Conventional DVC architectures, as usual, rely on a predefined GOP structure. For example, frames in a video are organized as I-WZ-I-WZ..., where the GOP size is 2. Figure 1 presents the architecture of a conventional transform-domain DVC scheme, where discrete cosine transform (DCT) is performed block-wisely before quantizing each WZF. Note that side information generated at the decoder is also DCT-transformed and then analyzed to provide the corresponding statistical properties for the channel decoder and the coefficient reconstructor. Then, WZFs are finally reconstructed after performing inverse DCT on reconstructed coefficients. Due to high correlation between

consecutive frames, WZFs are able to be well reconstructed if reliable side information is available. However, correlation between consecutive frames is often neither spatially nor temporally stationary. Hence, a fixed GOP structure will not be efficient enough in generating accurate side information all the time. This implies that the bit rates of required parity information for WZFs reconstruction will not be temporally smooth, or the frame reconstruction quality will not be even at a specified parity bit rate. Several works have been motivated to adapt the GOP structure to varying video content. Ascenso et al. [6] proposed a coding scheme with dynamic GOP structure according to four motion activity metrics (i.e., difference of histogram, histogram of difference, block histogram difference, and block variance difference) in a frame, by which each frame in a video sequence is thus classified into KF or WZF mode. Intuitively, for high-motion video, the distance (i.e., the GOP size) between two consecutive KFs should be shorter to keep the side information accurate, while a larger distance would be preferred for low-motion video. This flexible GOP structure can be considered as the temporal granularity in coding mode selection and is useful for DVC systems to remove temporal redundancy adaptively and obtain better performance compared to the architecture with a rigid GOP size.

2.2 Block-based DVC scheme

The frame-based WZ video codec with fixed GOP size assumes spatially stationary correlation noise statistics between a WZF and its side information generated by exploiting nearby KFs. This assumption however degrades accuracy of the side information for regions of unpredictable instances, like active motions and dis-occlusions, hence resulting in more parity bits required for decoding WZF.

In [23], the GOP coding structure is retained, but intra-coding is possibly performed for some highly dynamic

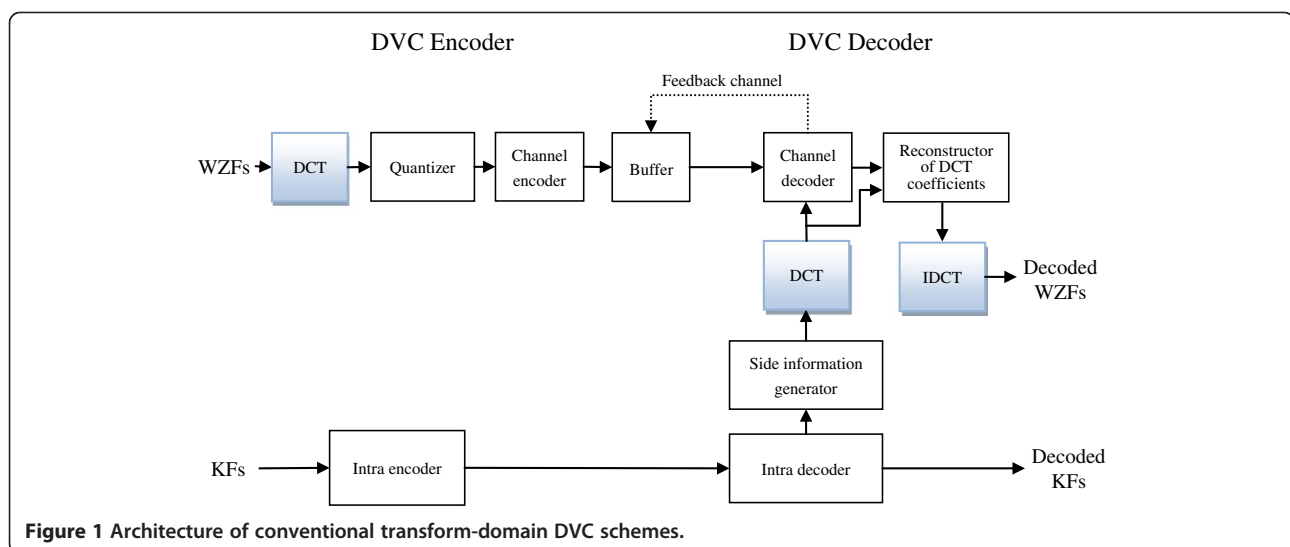


Figure 1 Architecture of conventional transform-domain DVC schemes.

blocks for non-key frames. Henceforth, block-based encoding with mixed modes (intra/WZ) is conducted in their work. On the other hand, Puri et al. [20] classify blocks by estimating their temporal correlation noises and enable different levels of channel coding accordingly. In contrast to traditional frame-based coding schemes, the above two methods perform WZ coding (based on, e.g., turbo code) locally on the block basis. These, though solving the problem of spatial non-stationarity, still retain the problem of low coding efficiency for static regions (e.g., backgrounds) in KFs. The mixed block-based intra/inter mode selection in traditional video coding standards motivates us to explore both temporal and spatial granularity to further improve the coding efficiency and flickering artifacts in traditional DVC systems.

3 Proposed framework

In conventional video coding techniques, the compression performance depends on how temporal and spatial redundancies are removed (i.e., how alike a block is with respect to a counterpart block in the reference frame). Similarly, WZ coding efficiency depends on correlation between each block and its side information. In view of this point, conventional DVC coding architecture is actually not suitable for providing higher WZ coding efficiency. The reasons are that most frame contents are only regionally dynamic or static (except those being incurred due to camera motion or scene change) and the side information is mostly derived from the previous frame in time. This leads to a locally varying correlation between each block and its side information within a WZ frame. According to this observation, a decision process is proposed here for the encoder to individually assign each block (instead of a full frame) to either WZ or intra-coding type based on certain spatial or temporal properties. In Table 1, several important symbols used in this paper are noted.

3.1 TGOB structure

The essentials of the proposed framework include spatio-temporal analysis in a block mode decision (BMD) process which identifies each 16×16 block in a frame as a KB or a WZB, depending on the test of two image features, called temporal difference TD and spatial variance σ^2 :

$$TD_{i,j,t} = \sum_{(m,n) \in B_{i,j,t}} |x_t(m,n) - x_{t-d}(m,n)|, \quad (1)$$

$$\sigma_{i,j,t}^2 = \frac{1}{M} \sum_{(m,n) \in B_{i,j,t}} |x_t(m,n)|^2 - \left(\frac{1}{M} \sum_{(m,n) \in B_{i,j,t}} x_t(m,n) \right)^2, \quad (2)$$

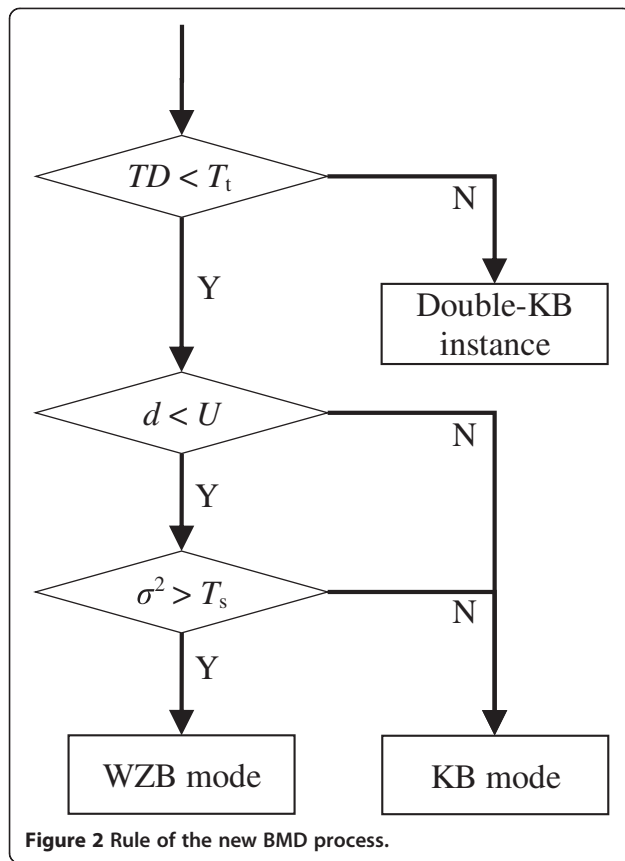
where (i, j) indicates the block location, t is the time index, (m, n) represents a pixel location in the current block $B_{i,j,t}$, x_t is the pixel value, M is the number of pixels within a

Table 1 Notations of symbols

Notation	Definition
U_{TGOB}	Predefined upper bound for the length of a TGOB
TD	Temporal difference between the current block and the preceding KB
T_t	Threshold of TD
σ^2	Spatial variance of a block
T_s	Threshold of σ^2
D	Temporal distance between the current block and the preceding KB
M_k	Number of bit planes selected for transmission of the DCT coefficients in the k th band
R	Residual between a WZB and its corresponding side information
A	Parameter of Laplacian distribution
QP	Quantization parameter for intra-coding
QM	Quantization matrix for WZ coding
D_k	Deviation of a coefficient from the k th band mean
μ_k	Mean of coefficients in the k th band for all WZBs of the current frame
σ_k^2	Variance of coefficients in the k th band for all WZBs of the current frame

block, and d is the temporal distance between the current block and the preceding KB. Two thresholds, T_t and T_s , are given for TD and σ^2 , respectively. Blocks with large TD or small σ^2 are identified as KBs for intra-coding, due to the observation that blocks with higher motion activity or higher spatial correlation are more suitable for intra-coding. Note that in our DVC system, side information is generated by interpolation from two bounding KBs. The accuracy of interpolation can be guaranteed if these two bounding KBs are similar in content. Hence, the block preceding (in time) the one satisfying the TD or σ^2 condition should be intra-coded, too. In view of this concept, three coding cases are designed in our scheme, including WZB mode, KB mode, and double-KB instance. The difference between the latter two cases is that two consecutive co-located blocks in time are assigned as KBs for the double-KB instance. Consecutive co-located WZBs in temporal direction, together with the preceding KB, form a TGOB. The TGOB size is at least 1 (i.e., containing only a leading KB and no WZBs) and large up to a predefined upper bound U_{TGOB} to avoid a long delay during side information generation at the decoder.

To ensure that side information for each WZB is interpolated from two similar KBs, a decision flow of block modes is proposed as in Figure 2. As illustrated in Figure 2, if the TD value for the current block is higher than T_t , the current block and its co-located one in the previous frame are both identified as KBs (denoted as KB2 and KB1, respectively, i.e., the double-KB instance).



Since a large motion or scene change occurs at KB2 (current block), KB1 (in the previous frame) will be paired with its preceding KB, while KB2 will be paired with another succeeding KB, both for side information generation. If TD is not higher than T_b , the current block might be WZ-coded. However, we have to check whether the TGOB size achieves the upper bound U_{TGOB} . If yes, a KB is assigned so that the TGOB size constraint is satisfied and a long delay is prevented. If not, a test on σ^2 is subsequently performed. For blocks whose variances σ^2 are small, they are identified as KBs, as mentioned earlier. Otherwise, a WZB mode is identified. Note that the WZB mode is assigned if both tests on TD and σ^2 in Figure 2 are satisfied and the length bound U_{TGOB} is not achieved. All the block coding modes (WZB or KB) are recorded as a block mode map (BMM) for each frame and transmitted to the decoder for reconstruction guidance.

KBs will be H.264/AVC intra-coded, while WZBs will be channel-encoded. Each WZB is decoded with the help of side information generated by bilinear interpolation of two co-located and reconstructed KBs that precede and succeed the current WZB, respectively. Different from our previous work [21], the double-KB instance is newly introduced in this new BMD process and aims at enhancing the side information quality for WZBs (i.e., ensuring that the

two KBs used for side information generation are similar in content). Note that the goal of the upper bound for TGOB size is to avoid a long delay during side information generation at the decoder. Consequently, the TGOB size will be small for dynamic regions and large for static content.

To better understand the TGOB structure, Figure 3 illustrates an example of dynamic TGOBs for a sequence of eight frames, where U_{TGOB} is set to be 4. In this sequence, the first and the last frames are all intra-coded. The (0, 1)th block in frame #3 is identified as KB2 in the double-KB instance, and the co-located block in frame #2 is forcedly changed to be a KB1 though it is initially identified as a WZB. Hence, the series of TGOB sizes at block position (0, 1) is {2, 1, 4, 1}; the side information for corresponding WZBs in frame #1 and frames (#4, #5, #6) are generated from pairs of KBs in frames (#0, #2) and (#3, #7), respectively.

Fine granularity in both the spatial and temporal domains provides a dynamic switch between WZBs and KBs, which offers better visual experience in terms of quality smoothness by reducing fluctuations between consecutive frames. This will be demonstrated in later sections.

3.2 TGOB-based DVC encoder

Figure 4 shows our DVC codec framework in accordance with the proposed TGOB structure. Two main components of the encoder are described as follows.

3.2.1 New BMD

The BMM determined based on the decision rules of Figure 2 is sent to the BMM queue at the DVC decoder. The load for BMM transmission is light with only 1,485 bits per second (QCIF format at a frame rate of 15 Hz and 16×16 block size). The identified KBs are, on one hand, sent for intra-coding and, on the other hand, replaced with zero blocks, which are then combined with the quantized DCT coefficients of WZBs for channel coding. Note that our DVC system, though performing BMD, still adopts frame-based LDPCA codes (see next subsection) to preserve its coding efficiency.

3.2.2 LDPCA encoder

A 4×4 DCT is applied to each WZB (note that KBs are replaced with zero blocks for channel encoding) to obtain DCT coefficients of varying digital frequencies. In a frame, transform coefficients of the same frequencies are grouped together to form 16 bands, each of which is encoded independently thereafter. Each DCT coefficient of the k th band is uniformly quantized to obtain a symbol q representing one of 2^{M_k} levels, where $k \in \{0, 1, \dots, 15\}$ is the band index ($k=0$ represents the DC band and $k=15$ represents the highest AC frequency) and $M_k \in \{0, 1, 2, 3, \dots, 8\}$ denotes the number of bit planes selected for transmission. M_k

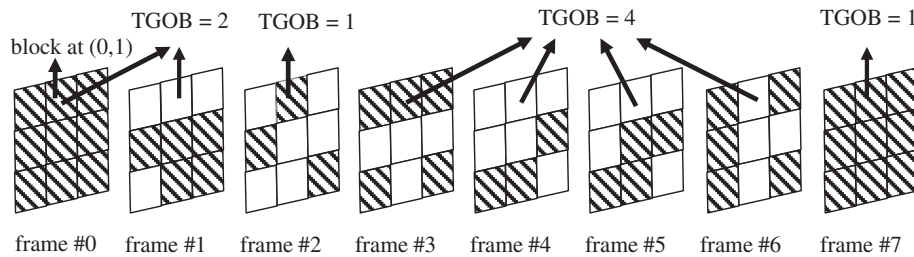


Figure 3 Illustration of dynamic TGOB structure with $U = 4$. Shaded blocks represent KBs, and the others are WZBs.

for each band is selected according to a predefined quantization matrix (QM) proposed in the DISCOVER codec [22], and each selected bit plane of each band is processed by the LDPCA encoder to produce parity bitstream which is then stored in an output buffer and partially sent on demand to the decoder in response to the feedback signal. The QMs are illustrated in Figure 5, where the number in each entry denotes the quantization level applied for each band. For example, in QM_8 , the DC band has 128 quantization levels and seven bit planes will be encoded as M_0 equals 7. For these matrices, QM_8 results in smaller quantization errors and higher bit rates, while QM_1 corresponds to larger quantization errors and lower bit rates. Note that though the number of identified WZBs in each frame is varying due to non-stationary video content, the filled zero blocks for KBs make the input to the LDPCA encoder a fixed size (i.e., full-frame size). This results in LDPCA inputs of regular size, at a cost of probable sacrifice in coding efficiency (i.e., more parity bits are required for transmission). Fortunately, this increased bit rate can be partly eliminated: the parity bits fully connected to 0 bits of the KBs could be discarded during the accumulation of parity bits in the LDPCA encoder. As shown in Figure 6, three parity bits

(dash squares in the second column) completely connected to the 0 bits from KBs are neither accumulated during encoding nor sent to the DVC decoder. Additionally, the inserted 0 bits will also yield a faster convergence of channel decoding due to the exactly correct side information at the decoder provided by the received BMMs (see next subsection). In our DVC system, the CRC bits for each bit plane are also multiplexed with the LDPCA bitstreams to verify the correctness of channel decoding convergence.

3.3 TGOB-based DVC decoder

At the DVC decoder, the intra-coded bitstream for KBs is first decoded and then sent to the decoded KB queue. With the received parity bits and the side information generated from the decoded KBs, DCT coefficients of WZBs can be decoded bit plane by bit plane by the LDPCA decoder. The reconstructor is used to reconstruct DCT coefficients from the decoded bit planes and de-group the bands before the 4×4 inverse DCT transform (IDCT) is applied. The reconstructed KBs and WZBs are finally combined, according to BMMs, to form consecutive frames of a complete video sequence. The details about DVC decoding are described as follows.

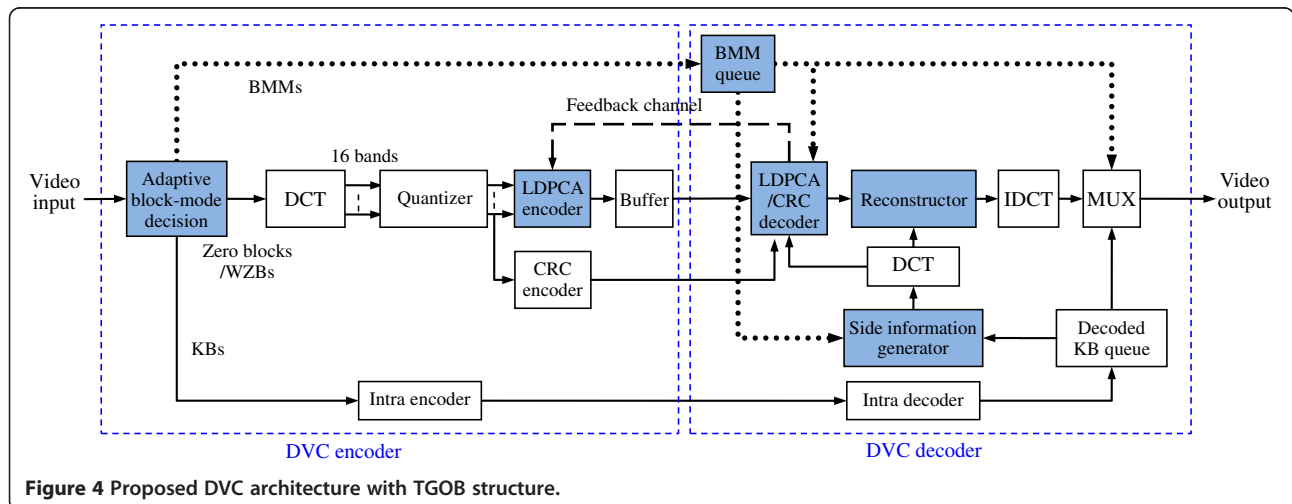
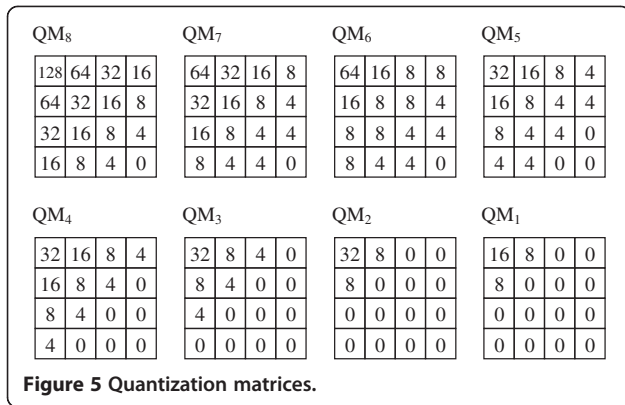


Figure 4 Proposed DVC architecture with TGOB structure.



3.3.1 BMM queue

BMM for each frame is received and stored in the queue, where it can be accessed by the LDPCA decoder and the multiplexer MUX to realize LDPCA decoding and to complete the whole frame reconstruction, respectively. Moreover, with the guidance of BMMs, each TGOB can be identified with a leading KB possibly followed by several WZBs (might be 0). Side information generation for WZBs within a TGOB can then be carried out based on the interpolation between the leading KB and the next KB that follows.

3.3.2 Side information generator

Against the traditional DVC decoder which is based on MCTI techniques [5,6] for side information generation, a simple bilinear interpolation [21] technique is adopted here to estimate WZBs based on two decoded KBs. Since each WZB needs its two bounding KBs for side information

generation, a buffer (the decoded KB queue in Figure 4) of $2 \cdot U_{TGOB} - 1$ frames is required at the decoder. When a WZB of the current frame is considered for reconstruction, the temporal positions of its two bounding KBs are known from the received BMM, by which decoded KB data can be accessed from the buffer and side information can be calculated. In our system, simple bilinear interpolation makes the complexity of the DVC decoder much reduced, thanks to the dynamic TGOB structure which ensures content similarity between two bounding KBs for any considered WZB.

3.3.3 CNM

In general, a good CNM can reduce the parity bit rate and achieve a higher compression ratio. Therefore, the establishment of an accurate CNM at the decoder is critical for the performance of a DVC system. Laplacian distribution is one of the most common CNMs adopted in DVC systems [4] to model the residual r between the original WZB and the corresponding side information block (SIB). In this paper, a similar Laplacian model is adopted for the statistics of each DCT coefficient c with mean μ , which is expressed as

$$p(c) = \frac{\alpha}{2} e^{-\alpha \cdot |c - \mu|}, \quad (3)$$

$$c(u, v) = \text{DCT}(r(x, y)) \text{ and}$$

$$r(x, y) = \text{WZB}(x, y) - \text{SIB}(x, y).$$

We need to estimate the coefficient-wise α at the decoder, without knowing the original WZB. An approximation is to estimate r as the half of the difference between two bounding KBs, KB^{front} and KB^{back} , i.e.,

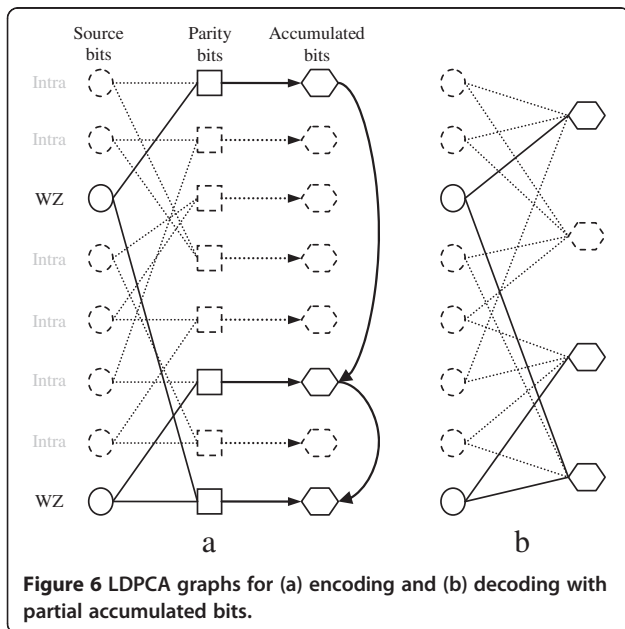
$$\hat{r}(x, y) = (\text{KB}^{\text{front}}(x, y) - \text{KB}^{\text{back}}(x, y)) / 2. \quad (4)$$

A 4×4 DCT is then applied to the estimated residual blocks $\hat{r}(x, y)$, where WZBs are identified, to obtain the estimated coefficients $\hat{t}(u, v)$. To estimate the Laplacian parameter $\hat{\alpha}$ from $\hat{c}(u, v)$ samples, the method in [4] is adopted and rewritten as below:

$$\hat{\alpha}_k(u, v) = \begin{cases} \sqrt{2/\sigma_k^2}, & [D_k(u, v)]^2 \leq \sigma_k^2 \\ \sqrt{\frac{2}{[D_k(u, v)]^2}}, & [D_k(u, v)]^2 > \sigma_k^2 \end{cases}, \quad (5)$$

$$D_k(u, v) = |\hat{c}(u, v) - \mu_k|, \quad \hat{c}(u, v) \in k\text{th band} \quad (6)$$

where μ_k and σ_k^2 are mean and variance of $|\hat{c}(u, v)|$'s in the k th band for all WZBs of the current frame. $D_k(u, v)$ measures the deviation of the coefficient $\hat{c}(u, v)$ from the corresponding band mean. If $[D_k(u, v)]^2 \leq \sigma_k^2$, $\hat{c}(u, v)$



is considered as an inlier coefficient and σ_k^2 is used to compute $\hat{\alpha}$. Otherwise (i.e., $\hat{c}(u, v)$ is an outlier), $[D_k(u, v)]^2$ is used instead.

3.3.4 LDPCA decoder

Based on the requested parity bits transmitted from the encoder and the received BMM, each bit plane can be decoded with reference to the side information and the CNM characterizing the statistical dependency between WZBs and corresponding SIBs. This decoding process considers not only the parity bits, the temporal characteristics, and the already decoded bits in the more significant bit planes, but also those 0 bits contributed by KBs and indicated by BMM. Note that more KBs in a frame actually lead to more efficient LDPCA decoding, thanks to the reliable side information contributed by KBs. As mentioned in Figure 6a, parity bits fully connected to the 0 bits of KBs are neither accumulated nor transmitted. Hence, to realize LDPCA decoding with a regular size, these non-transmitted bits should be recovered by inserting 0's into right positions according to the received BMM, as illustrated by the dashed hexagon in Figure 6b. Moreover, transmitted parity bits which are partly connected to 0 bits of KBs are helpful for fast convergence in decoding. For example, the first accumulated parity bit, denoted as the first hexagon shown in the second column of Figure 6b, is connected to one WZB bit and three KB bits. Due to reliable 0 bits from KBs, determination of the WZB bits becomes easier during message passing iterations. Finally, CRC bits are used to confirm the convergence of LDPCA decoding.

3.3.5 Reconstruction

After the bit planes of the quantized DCT coefficients are decoded, the statistical CNM (estimated by the techniques mentioned in Section 3.3.2, i.e., $\hat{\alpha}$), the corresponding interval $[l, u)$ (indicated by the decoded quantization level q'), and the side information x' , used in the LDPCA decoding, are also exploited for optimal reconstruction of DCT coefficients in the sense of minimum mean square errors (MMSE) [24]. Each DCT coefficient in a WZB can be optimally reconstructed as \hat{x}_{opt} by Equation 7:

$$\hat{x}_{opt} = E[x|q', x'] = \begin{cases} l + \frac{1}{\hat{\alpha}} + \frac{u-l}{1-e^{\hat{\alpha}(u-l)}}, & x' < l \\ x' + \frac{\left(x' - l + \frac{1}{\hat{\alpha}}\right)e^{-\hat{\alpha}(x'-l)} - \left(u - x' + \frac{1}{\hat{\alpha}}\right)e^{-\hat{\alpha}(u-x')}}{2 - e^{-\hat{\alpha}(x'-l)} - e^{-\hat{\alpha}(u-x')}}}, & x' \in [l, u) \\ u - \frac{1}{\hat{\alpha}} - \frac{u-l}{1-e^{\hat{\alpha}(u-l)}}, & x' \geq u \end{cases}$$

Table 2 Configurations of GOP size, U_{TGOB} , and (QM, QP) quantization parameter pairs for DISCOVER and proposed schemes

Sequence	Foreman	Hall monitor	Soccer	Coastguard
GOP	2	8	2	2
U_{TGOB}	9	10	2	2
QM ₁				
QP (DISCOVER)	40	37	44	38
QP (proposed)	35	29	33	34
QM ₂				
QP (DISCOVER)	39	36	43	37
QP (proposed)	32	28	31	30
QM ₃				
QP (DISCOVER)	38	36	41	37
QP (proposed)	31	28	31	30
QM ₄				
QP (DISCOVER)	34	33	36	34
QP (proposed)	30	27	30	27
QM ₅				
QP (DISCOVER)	34	33	36	33
QP (proposed)	30	27	30	27
QM ₆				
QP (DISCOVER)	32	31	34	31
QP (proposed)	29	26	29	26
QM ₇				
QP (DISCOVER)	29	29	31	30
QP (proposed)	26	23	27	26
QM ₈				
QP (DISCOVER)	25	24	25	26
QP (proposed)	25	21	25	25

4 Simulation results

Four test sequences, including 'Foreman', 'Hall monitor', 'Soccer', and 'Coastguard' of QCIF format at 15 Hz, are used to evaluate the performance of the proposed DVC scheme in terms of R-D behavior and temporal flicker after reconstruction at the decoder. The experiments were

performed on a PC platform with an i7-2600 3.40-GHz CPU and 6 GB RAM. The classified KBs are intra-coded by H.264/AVC reference software (JM15.1). The bit rates and the peak signal-to-noise ratios (PSNRs) are averaged over the whole sequence including KBs and WZBs. Parameters in our proposed WZ coding scheme include the following: upper bound (U_{TGOB}) of the TGOB size and thresholds for TD in Equation 1 and σ^2 in Equation 2, namely T_t and T_s . Values of T_t and T_s were experimentally obtained and set to 1,000 and 10, respectively. An increase of T_t or a decrease of T_s will cause a decrease of KB percentage (or an increase of WZ percentage), which would be more suitable for sequences of stationary and simple content (e.g., indoor video surveillance applications). To make the performance comparisons fairer, the GOP size (2 ~ 10) for the DISCOVER system is determined by optimizing the coding efficiency for each individual sequence. Similarly, our proposed scheme also optimizes U_{TGOB} (2 ~ 10) of the TGOB size for each individual sequence. Configurations of QPs for KFs/KBs at specified QMs for WZFs/WZBs to achieve variable bit rates and values of selected GOP size and U_{TGOB} are listed in Table 2 for comparison between DISCOVER and proposed schemes.

Table 3 Quality gains (in BDPSNR) for the proposed DVC scheme with new BMD against other DVC schemes

	Foreman (dB)	Hall monitor (dB)	Soccer (dB)	Coastguard (dB)
DISCOVER	0.61	2.85	2.19	-1.45
Proposed with original BMD	-0.04	-0.06	0.16	-0.03
H.264/AVC Intra	0.55	7.18	0.04	-0.28

4.1 Comparison on R-D performance

R-D performances of several coding schemes are compared in Figure 7. We apply the Bjøntegaard delta metric (BJM) to reveal the quality gain [25] against the DISCOVER system [22], which is known to use a frame-based transform-domain DVC scheme. Table 3 shows that our proposed scheme has Bjøntegaard delta PSNR (BDPSNR) gains by 0.61, 2.85, 2.19, and -1.45 dB for the sequences Foreman, Hall monitor, Soccer, and Coastguard, respectively. Note that the higher the temporal dependency is (e.g., the static sequence Hall monitor), the more the quality gain we have. Moreover, our scheme also outperforms 'H.264/AVC Intra' (equivalent to 'Proposed U1' and 'DISCOVER GOP1') for

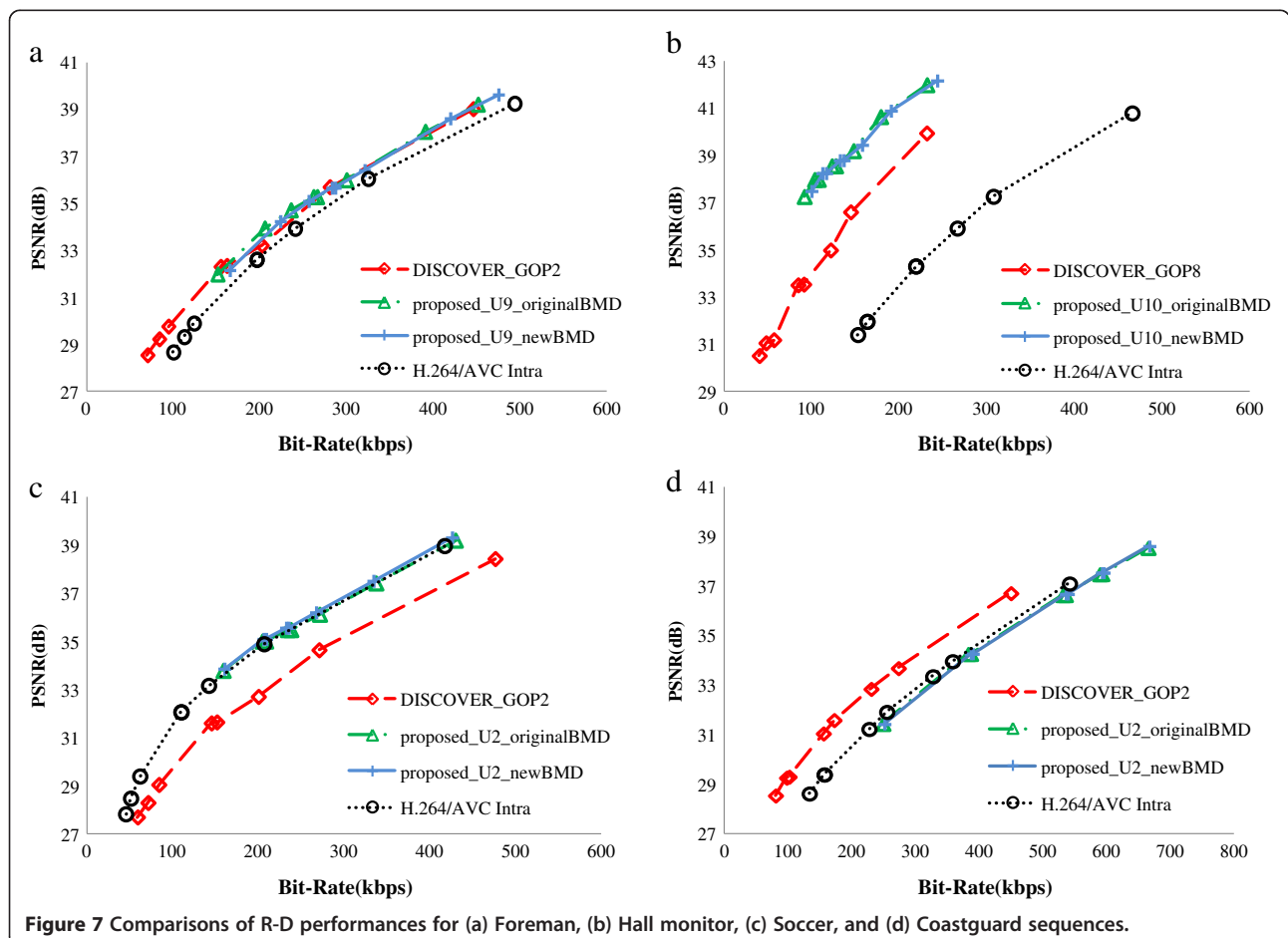


Figure 7 Comparisons of R-D performances for (a) Foreman, (b) Hall monitor, (c) Soccer, and (d) Coastguard sequences.



Figure 8 Reconstructed frame #17 in Hall monitor with (a) DISCOVER scheme and (b) the proposed one.

all sequences except Coastguard. It is found that our DVC scheme is not beneficial to the encoding of highly dynamic video, where most of the image blocks are classified as KBs (see the next paragraph for statistics). Fortunately, a highly dynamic video is rare in realistic DVC applications.

The percentages of KBs for the proposed scheme are found to be 91.11% (Foreman), 29.77% (Hall monitor), 96.57% (Soccer), and 98.17% (Coastguard), which are higher, by 10.35%, 1.93%, 4.12%, and 3.38%, than when the original BMD [21] (without double-KB instance) is adopted. This increase in KB percentages, though possibly increasing the bit rates, results in a comparable BDPSNR performance (see Table 3), all thanks to the accurate and improved side information generation by forcedly inserting KB1. Combining the above KB percentage statistics, QM/QP configurations in Table 2, and PSNR figures in Table 3, it is deduced that intra-coded KBs do not necessarily enhance or degrade the coding efficiency, but the proper mode switch that determines the accuracy of side information and thus the coding efficiency.

Later, we will show that temporal flickering artifacts are further reduced with the new BMD proposed in the current system. For visual quality, the appearance of the passerby in Figure 8 obtained based on the proposed scheme is much clearer than that obtained based on DISCOVER, especially around the moving object boundaries.

4.2 Comparison on temporal flicker

4.2.1 Temporal flickering reduction by enlarging GOP size

Temporal smoothness is a key factor for the human visual system to evaluate the visual quality during the playback of

a video sequence. Generally, a decoded video sequence with lower flicker is more preferable, whether for conventional video coding schemes or DVC schemes. In particular, a DVC scheme will suffer from quality unbalance and thus annoying visual perception between KFs and WZFs if the quantization parameters are not well chosen. The DISCOVER system suggested eight quantization parameter pairs (QM, QP) for WZFs and KFs according to heuristic experiments which target at maximizing the R-D performance while keeping image qualities of the WZFs and KFs similar. Even though the reconstructed WZFs and KFs have similar PSNRs, it is not guaranteed to have low flicker while playing the decoded video sequences. Intuitively, flickering artifacts can be reduced if the coding mode switch is kept less frequent [26]. This can normally be accomplished by choosing a larger GOP size.

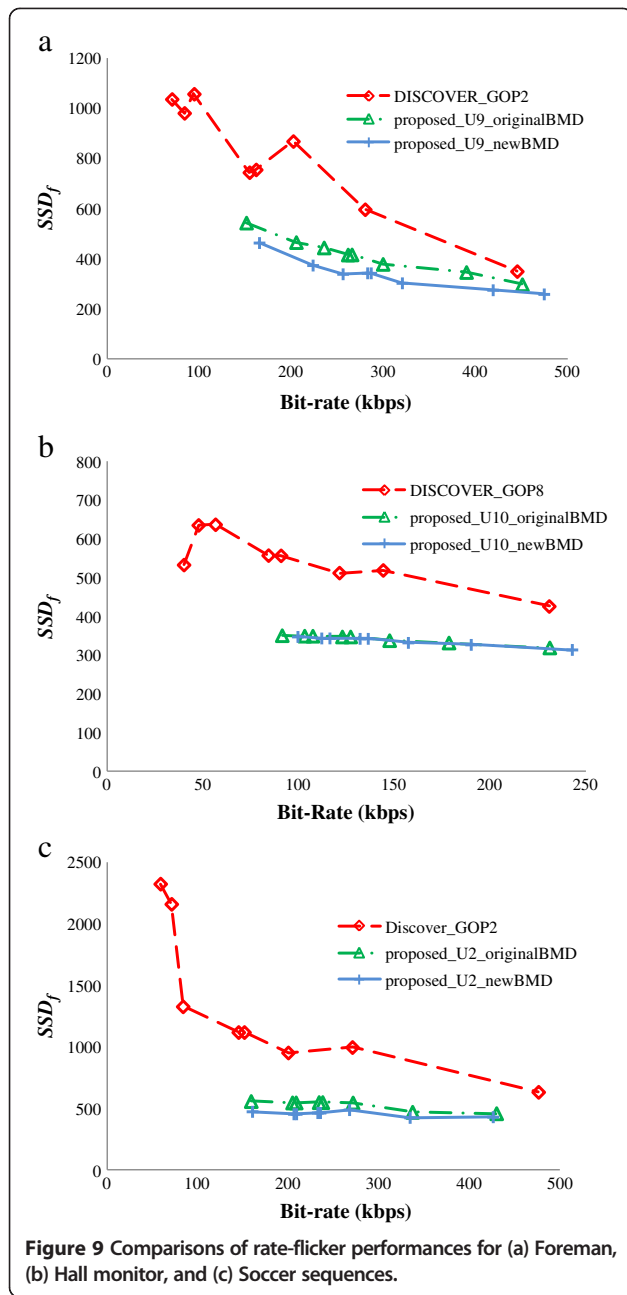
The sum of squared difference (SSD) metric proposed in [27] can fairly evaluate the temporal fluctuation of the decoded video sequence, as recalled in Equation 8:

$$\begin{aligned}
 SSD_f &= \frac{1}{\text{card}(I)} \sum_{(x,t) \in I_i} [e_t(x) - e'_t(x)]^2 \\
 I &= \left\{ I_i \mid \sum_{x \in I_i} [e_t(x)]^2 < \tau, \quad 1 \leq i \leq N \right\}, \\
 e_t(x) &= |f_t(x) - f_{t-1}(x)| \\
 e'_t(x) &= |f'_t(x) - f'_{t-1}(x)|
 \end{aligned} \tag{8}$$

where $f_t(x)$ and $f_{t-1}(x)$ represent the values of the x th pixel in the t th and $(t-1)$ th frame of the original video sequence, respectively, $e_t(x)$ represents the difference of

Table 4 Flickering reduction for Hall monitor encoded by DISCOVER with a larger GOP size

	QM ₁	QM ₂	QM ₃	QM ₄	QM ₅	QM ₆	QM ₇	QM ₈
GOP2	1,557	1,467	1,470	1,078	1,078	880	812	558
GOP8	533	636	637	557	557	512	519	426
Flickering reduction	66%	57%	57%	48%	48%	42%	36%	24%



co-located pixels between adjacent frames, and $f'_i(x)$, $f'_{i-1}(x)$, and $e'_i(x)$ have similar meanings, but defined for the reconstructed video sequence. I_i denotes the i th stationary block satisfying low-motion check with a threshold τ , N denotes the number of static blocks in a frame, $\text{card}(I)$ represents the cardinality of the set I , and SSD_f stands for the flickering intensity between $f_i(x)$ and $f_{i-1}(x)$. Note that blocks of higher motion activity are excluded for flickering calculation by choosing $\tau = 500$ (as suggested in [28]) in our system.

To verify the influence of the GOP size on temporal fluctuation, the sequence Hall monitor is encoded by the DISCOVER codec with two GOP sizes, namely 8 and 2. Table 4 lists the flickering values of the reconstructed video, showing that up to 66% of flickering reduction can be achieved for low-quality video encoded with QM_1 and GOP8.

4.2.2 Performance comparison on temporal flicker

In Figure 9, the rate-flicker performances for (QM, QP) pairs configured as in Table 2 are measured according to the SSD_f definition in Equation 8. The flickering reduction is defined as

$$\Delta\text{SSD} = \frac{\text{SSD}' - \text{SSD}}{\text{SSD}'} \times 100\%, \quad (9)$$

where SSD is the flickering value at a given bit rate for the proposed scheme and SSD' is the flickering value for DISCOVER which is interpolated/extrapolated at the same bit rate. Flickering evaluation does not apply to Coastguard since it contains highly dynamic content and none of the blocks satisfies the low-motion check in Equation 8. Tables 5, 6, and 7 show the flickering reduction ΔSSD for the rest of the sequences. Our proposed DVC scheme with new BMD obviously outperforms DISCOVER, where the flickering reductions in average are 39%, 33%, and 51% for Foreman, Hall monitor, and Soccer, respectively. Besides, further flickering reduction is also verified for the new BMD with respect to the original BMD.

Our proposed DVC scheme presents fine coding mode (KB or WZB) granularity in both the temporal and spatial domains, hence resulting in a highly temporal smoothness, compared to the conventional frame-based schemes. In

Table 5 Comparisons on measures of flickering intensities for Foreman against DISCOVER at eight given bit rates

	Proposed DVC with original BMD								Proposed DVC with new BMD							
	150	205	235	261	265	299	390	450	165	223	255	282	286	320	419	474
SSD (proposed)	543	465	444	416	416	378	346	297	463	374	339	343	343	303	275	259
SSD' (DISCOVER)	762	857	752	661	646	567	432	341	765	794	680	592	587	536	389	306
ΔSSD (%)	29	46	41	37	36	33	20	13	39	53	50	42	42	44	29	15
Average (%)	32%								39%							

Table 6 Comparisons on measures of flickering intensities for Hall monitor against DISCOVER at eight given bit rates

Bit rate	Original BMD								New BMD							
	91	103	107	123	127	147	178	231	99	112	116	132	136	157	190	243
SSD (proposed)	350	349	349	347	347	338	331	319	347	343	343	342	342	332	327	314
SSD' (DISCOVER)	556	539	533	513	514	516	482	426	544	526	520	515	517	505	470	413
Δ SSD (%)	37	35	34	32	32	35	31	25	36	35	34	34	34	34	30	24
Average (%)	33%								33%							

particular, the dynamic TGOB structure is capable of adapting the DVC system to varying video content so that the KB/WZB mode switch only occurs at regions of scene change or high motion. Our new BMD, introducing the double-KB instance, ensures that two bounding KBs of each WZB are similar in content and thus is seen to have an improvement against the original BMD in terms of flickering reduction.

4.3 Complexity analysis for DVC encoder

It is obvious that our proposed block-based DVC scheme is beneficial in the reduction of decoding complexity since MCTI at both encoder and decoder sides can be avoided. The overhead of BMD that brings this benefit needs to be further analyzed at the encoder. First, we measure the time needed to perform intra-encoding, WZ encoding, and BMD, all for full-frame unit, and denote them as α , β , and γ , respectively. Assume that the average percentage of KBs in a frame is ρ . The encoding time for a frame for our proposed scheme can be expressed as

$$T_{proposed} = \rho \cdot \alpha + \beta + \gamma. \quad (10)$$

Note that full-frame LDPCA coding is performed even if our scheme is block-based. On the other hand, for the conventional frame-based DVC scheme, the average encoding time for a frame is estimated to be

$$T_{conventional} = \begin{cases} \alpha & \text{key frame} \\ \beta & \text{WZ frame} \end{cases} \text{ or } T_{conventional}^{average} = 0.5 \cdot \alpha + 0.5 \beta. \quad (11)$$

Note that Equation 11 is estimated by assuming a GOP size of 2. According to experiments on our PC platform, WZ encoding actually has a lower complexity

than intra-coding ($\beta = 0.2647\alpha$) and the overhead caused by BMD is extremely low ($\gamma = 0.0023\alpha$). From Equation 11, it is obvious that there is a significant fluctuation (switching between α and β) among frames in computing load for the conventional frame-based system. Contrarily, load fluctuation of our proposed system depends on the variation of KB percentage ρ , which is small compared to $|\alpha - \beta|/\alpha \approx 0.7353$ associated with the traditional DVC encoders. Loading fluctuation between two successive frames incurs difficulties in designing a stable frame-based DVC encoder.

For the average computing load, our scheme is measured at $\rho \cdot \alpha + \beta + \gamma = (0.2670 + \rho) \cdot \alpha$ for one frame, while the conventional DVC encoder is measured at $0.5\alpha + 0.5\beta = 0.6324\alpha$. That is to say, our scheme will have a lower average complexity when the KB percentage ρ is below 36.54%. Recalling from the prior statement in Section 4.1 where $\rho = 91.11\%$, 29.77%, 96.57%, and 98.17% for the four test sequences, our scheme has a lower computing complexity for the Hall monitor. For the other three test sequences, our scheme, though having a higher complexity, is beneficial in less loading fluctuation. Theoretically, to lower down the complexity, the two thresholds T_t and T_s of our scheme can be further fine-tuned.

5 Conclusions

The proposed DVC scheme based on a dynamic TGOB structure obviously adapts our system to varying video content and achieves less flickering artifact while maintaining high coding performance. The encoder in the proposed DVC system assesses the spatial-temporal properties (i.e., TD and σ^2 in Equations 1 and 2) of each image block to dynamically determine suitable modes (KB/WZB) for encoding. This fine granular, accurate, and dynamic mode

Table 7 Comparisons on measures of flickering intensities for Soccer against DISCOVER at eight given bit rates

Bit rate	Original BMD								New BMD							
	159	204	208	234	238	271	337	430	160	206	209	232	235	267	334	426
SSD (proposed)	560	548	548	551	548	472	461	319	477	457	457	467	467	491	428	433
SSD' (DISCOVER)	1,095	957	960	975	978	997	881	717	1,090	958	960	974	976	996	886	723
Δ SSD (%)	49	43	43	43	44	45	46	36	56	52	52	52	52	51	52	40
Average (%)	44%								51%							

switch based on blocks also simplifies side information generation at the decoder, resulting in a simple bilinear interpolation from two bounding KBs for each WZB. Further analyses reveal that our block mode decision, though might increase the computing load slightly, actually results in a less loading fluctuation along the encoding process. Experimental results show that R-D performance is enhanced and temporal flickering artifacts are reduced significantly, compared to the well-known DISCOVER system. However, our DVC design is not suitable for videos of high dynamics like Coastguard. In that case, all intra-coding could be conducted instead.

To further advance the R-D performance of this TGOB structure, block mode decision based on R-D optimization, instead of the threshold-based rule, can be considered. This R-D optimization, however, should be simplified and effective so that encoder complexity will not be increased too much. A research based on three or more modes (i.e., more than KB and WZB) is also ongoing, aiming to further enhance the R-D and flickering performances.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Science Council, Republic of China, under contracts NSC-101-2221-E-194-034 and NSC98-2221-E-194-037-MY3.

Author details

¹Department of Communications Engineering, National Chung Cheng University, Chia-Yi 621, Taiwan. ²Department of Electrical Engineering, National Chung Cheng University, Chia-Yi 621, Taiwan. ³Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI), National Chung Cheng University, Chia-Yi 621, Taiwan.

Received: 20 July 2013 Accepted: 13 November 2013

Published: 1 December 2013

References

1. B Girod, A Aaron, S Rane, D Rebollo-Monedero, Distributed video coding. *Proc. IEEE, Special issue on advance in video coding and delivery* **93**(1), 71–83 (2005)
2. JD Slepian, JK Wolf, Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory* **IT-19**(4), 471–480 (1973)
3. D Wyner, J Ziv, The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inform. Theory* **22**(1), 1–10 (1976)
4. C Brites, F Pereira, Correlation noise modeling for efficient pixel and transform domain Wyner–Ziv video coding. *IEEE Trans. Circuits Syst. Video Technol.* **18**(9), 1177–1190 (2008)
5. J Ascenso, C Brites, F Pereira, Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding, in *Proc. of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services* (Smolenice, Slovak Republic, 2005)
6. J Ascenso, C Brites, F Pereira, Content adaptive Wyner–Ziv video coding driven by motion activity. *Proc. Int. Conf. Image Process.* **20**, 605–608 (2006)
7. A Abou-Elailah, F Dufaux, J Farah, M Cagnazzo, B Pesquet-Popescu, Fusion of global and local motion estimation for distributed video coding. *IEEE Trans. Circuits Syst. Video Technol.* **20**(1), 158–172 (2013)
8. C Brites, J Ascenso, F Pereira, Side information creation for efficient Wyner–Ziv video coding: classifying and reviewing. *Signal Process. Image Commun.* **28**, 689–726 (2013)
9. G Petrazzuoli, M Cagnazzo, B Pesquet-Popescu, Novel solutions for side information generation and fusion in multiview DVC. *EURASIP J. Adv. Signal Process.* **2013**, 154 (2013)
10. FS-Y Chien, T-Y Cheng, S-H Ou, C-C Chiu, C-H Lee, V Srinivasa Somayazulu, Y-K Chen, Power consumption analysis for distributed video sensors in machine-to-machine networks. *IEEE J. Emerg. Select Top. Circuits Syst.* **3**(1), 55–64 (2013)
11. G-F Javier, Compression of correlated binary sources using turbo codes. *IEEE Comm. Lett.* **5**(10), 417–419 (2001)
12. A Aaron, B Girod, Compression with side information using turbo codes, in *Proc. of IEEE Data Compression Conference* (Snowbird, UT, 2002)
13. A Liveris, Z Xiong, C Georghiades, Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Comm. Lett.* **6**(10), 440–442 (2004)
14. D Varodayan, A Aaron, B Girod, Rate-adaptive codes for distributed source coding. *EURASIP Signal Process.* **86**, 3123–3130 (2006)
15. J Škorupa, J Slowack, S Mys, N Deligiannis, J De Cock, P Lambert, C Grecos, A Munteanu, R Van de Walle, Efficient low-delay distributed video coding. *IEEE Trans. Circuits Syst. Video Technol.* **22**(4), 530–544 (2012)
16. M Badem, WAC Fernando, AM Kondo, Unidirectional distributed video coding using dynamic parity allocation and improved reconstruction, in *Proc. of IEEE Int'l Conf. on Information and Automation for Sustainability (ICIAFs)* (Colombo, 2010)
17. D Kubasov, K Lajnef, C Guillemot, A hybrid encoder/decoder rate control for a Wyner–Ziv video codec with a feedback channel, in *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing* (Crete, 2007)
18. X Artigas, J Ascenso, M Dalai, S Klomp, D Kubasov, M Ouaret, The DISCOVER codec: Architecture, techniques and evaluation, in *Proc. of Picture Coding Symposium* (Lisbon, 2007)
19. A Aaron, S Rane, E Setton, B Girod, Transform-domain Wyner–Ziv codec for video, in *Proc. of The Visual Communications and Image Processing Conference* (San Jose, 2004)
20. R Puri, A Majumdar, K Ramchandran, PRISM: A video coding paradigm with motion estimation at the decoder. *IEEE Trans. Image Process.* **16**(10), 2436–2448 (2007)
21. D-C Tsai, C-M Lee, W-N Lie, Dynamic key block decision with spatio-temporal analysis for Wyner–Ziv video coding, in *Proc. of IEEE Int'l Conf. on Image Processing (ICIP)* (San Antonio, TX, 2007)
22. DISCOVER official site. http://www.discoverdvc.org/cont_Codec.html. Accessed on July 14, 2013
23. M Tagliasacchi, A Trapanese, S Tubaro, J Ascenso, C Brites, F Pereira, Intra mode decision based on spatio-temporal cues in pixel domain Wyner–Ziv video coding, in *Proc. of IEEE Int'l Conf. Acoustics, Speech, and Signal Processing* (Toulouse, 2006), pp. 57–60
24. J Nayak, C Guillemot, Optimal reconstruction in Wyner–Ziv video coding with multiple side information, in *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing* (Crete, 2007), pp. 183–186
25. G Bjontegaard, Calculation of average PSNR differences between RD-curves, in *VCEG Contribution VCEG-M33* (Texas, 2001)
26. JX Yang, HR Wu, Robust filtering technique for reduction of temporal fluctuation in H.264 video sequences. *IEEE Trans. Circuits Syst. Video Technol.* **20**(39), 458–462 (2010)
27. X Fan, W Gao, Y Lu, D Zhao, Flicking reduction in all intra frame coding, in *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q6, JVT 5th Meeting* (Geneva, 2002)
28. SS Chun, J-R Kim, S Sull, Intra prediction mode selection for flicker reduction in H.264/AVC. *IEEE Trans. Consum. Electron.* **52**(4), 1303–1310 (2006)

doi:10.1186/1687-6180-2013-177

Cite this article as: Lee et al.: Distributed video coding with block mode decision to reduce temporal flickering. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:177.